Journal of Statistical Research *ISSN 0256 - 422 X*
2007, Vol. 41, No. 2, pp. 63-71
Bangladesh

A CLASS OF ESTIMATORS OF REGRESSION COEFFICIENT FOR SIGN CHANGE PROBLEM IN MEASUREMENT ERROR MODELS

Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur-208016, India
Email: shalab@iitk.ac.in, shalabh1@yahoo.com

Alan T.K. Wan
Department of Management Sciences
City University of Hong Kong, Kowloon, Hong Kong
Email: msawan@cityu.edu.hk

summary

Srivastava and Shalabh (1997a, *Journal of Econometrics*) proposed a class of Stein-like consistent estimators for estimating the slope coefficient in a single explanatory variable ultrastructural model. This paper studies the sign reversal problem of this class of estimators and proposes an alternative class of improved estimators along the lines of the double-k class estimators of Ullah and Ullah (1978, *Econometrica*) that overcomes this problem. Large sample asymptotic properties of the proposed estimators are studied for the case where the distributions of the measurement errors are not necessarily normal.

Keywords and phrases: Consistency, large sample asymptotic, measurement errors, sign change, Stein-rule, ultrastructural models

AMS Classification: 62J05

1 Introduction

Stein-rule estimators proposed by Stein (1956) are well known for their properties of reducing the mean squared error of estimators of regression coefficients under some mild constraints on the shrinkage parameter provided that the number of coefficients is not less than three, see Saleh (2006) for an annotated bibliography on Stein-rule estimation. In the context of measurement error models, the Stein-rule procedure annihilates the inconsistency property of the least squares method and provides a class of consistent estimators (Srivastava and Shalabh (1997a, b)). Such estimators, however, suffer from the limitation of sign change, *i.e.*, the sign of the Stein-rule estimate may be opposite to that of the corresponding least square estimate. This can be circumvented by using a positive-part Stein-rule procedure which assigns the value zero to the parameter if sign change occurs. Apart from being an ad-hoc procedure, the positive-part Stein-rule estimator has the disadvantage of having rather intricate sampling properties, from which it is hard to draw any neat inference related to the efficiency properties of estimators even in large samples when the variables are contaminated with measurement errors. This paper proposes a simple way to avert the problem of sign change in the estimates along the lines of the double-k class estimators proposed by Ullah and Ullah (1978) for the linear regression model without measurement errors. We consider the ultrastructural formulation of a measurement error model involving a single explanatory variable and derive the large-sample asymptotic properties of the proposed estimator without assuming a normal distribution for the measurement errors.

2 The ultrastructural model and estimators

Let us postulate the following linear relationship

$$Y_i = \alpha + \beta X_i \quad ; \quad i = 1, 2, \ldots, n \tag{2.1}$$

where Y_i and X_i denote the ith observation of the study and explanatory variables respectively, α is the intercept term and β is the slope coefficient. It is assumed that both the variables are subject to measurement errors so that instead of observing the true values Y_i and X_i, we observe y_i and x_i, respectively such that

$$y_i = Y_i + u_i \tag{2.2}$$

and

$$x_i = X_i + v_i \quad , \quad i = 1, 2, \ldots, n \tag{2.3}$$

where u_i and v_i are the measurement errors, which are assumed to be i.i.d. random variables with zero mean and a common variance, *i.e.*, $u_i \sim i.i.d.(0, \sigma_u^2)$ and $v_i \sim i.i.d.(0, \sigma_v^2)$. The conventional disturbance term in the regression relationship is assumed to be subsumed in u_i without any loss of loss of generality. Further, the true values X_i's are assumed to be have means m_i's such that,

$$X_i = m_i + w_i \quad i = 1, 2, \ldots, n \tag{2.4}$$

where $w_i \sim i.i.d.(0, \sigma_w^2)$. If X_i's are fixed and $w_i = 0$ for all i's, then it becomes the functional form of the measurement error models. On the other hand, if X_i's are stochastic but m_i's are identical, then the ultrastructural form reduces to the structural form (Dolby (1976)). Finally, we assume that the error terms $u_i, v_i$and $w_i$are mutually independent but follow distributions which are not necessarily normal.

Now, rewriting (2.1) as,

$$y_i = \alpha + \beta x_i + (u_i - \beta v_i) \quad , \quad i = 1, 2, \ldots, n. \tag{2.5}$$

The least squares estimator of β is simply

$$b = \frac{s_{xy}}{s_{xx}} \ , \tag{2.6}$$

where

$$\begin{aligned} s_{xx} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \\ s_{xy} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \\ \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \end{aligned}$$

and

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

It is well documented that b is both a biased and an inconsistent estimator of β. under the model (2.1)-(2.4), see Fuller (1987, Ch.1) and Cheng and Van Ness (1999, Ch.1). Suppose, however, that σ_v^2, the variance of the measurement errors associated with the explanatory variable, is known a priori, then a consistent estimator of β may be formulated as

$$\hat{\beta} = \frac{s_{xy}}{s_{xx} - \sigma_v^2} \quad ; \quad s_{xx} > \sigma_v^2, \tag{2.7}$$

which is commonly known as the adjusted or immaculate estimator. Note that $\hat{\beta}$ is a consistent estimator of β (Schneeweiss (1976), Fuller (1987, Ch. 1). In the case of measurement error models with replicated observations, consistent estimators have been obtained and analyzed in Shalabh (2003).

As an alternative to the immaculate estimator, Srivastava and Shalabh (1997a, b) suggested a utilization of the Stein-rule procedure (Stein (1956)) with Lindley-like mean correction on X_i's, leading to,

$$\hat{X}_i = \bar{x} + \left[1 - \left(1 - \frac{g}{n}\right)\frac{\sigma_v^2}{s_{xx}}\right](x_i - \bar{x}), \tag{2.8}$$

where g is a non-negative scalar. See also Shalabh (1998, 2000) for an alternative application of the Stein-rule procedure in models involving several explanatory variables. Note that $E(\hat{X}_i - X_i)$ generally differs from zero. If one uses $\hat{X}_i$ in place of X_i in (2.1), after replacing Y_i by $(y_i - u_i)$ and applies least squares, then the following g-class consistent estimator of β is obtained:

$$\hat{\beta}_g = \left[1 - \left(1 - \frac{g}{n}\right)\frac{\sigma_v^2}{s_{xx}}\right]^{-1} b. \tag{2.9}$$

Srivastava and Shalabh (1997a) studied the large sample properties of the g-class estimator and derived conditions for $\hat{\beta}_g$ to dominate $\hat{\beta}$ in mean squared error (MSE) criterion. In particular, it is found that the reduction in MSE is maximized when g is taken as

$$g \equiv g_d = r + \frac{3(2+\theta\gamma_{2v})}{(1-\theta)} \tag{2.10}$$

where

$$\begin{aligned}
\theta &= \frac{\sigma_v^2}{s_{mm}+\sigma_w^2+\sigma_v^2}, \quad 0 \le \theta < 1 \\
s_{mm} &= \frac{1}{n}\sum_{i=1}^{n}(m_i-\bar{m})^2, \\
\bar{m} &= \frac{1}{n}\sum_{i=1}^{n} m_i, \\
r &= \frac{\sigma_u^2}{(1-\theta)\beta^2\sigma_v^2}
\end{aligned}$$

and γ_{2v} denotes the Pearson's measure of the excess kurtosis of the distribution of the v_i's. Note that $\theta = 0$ when $\sigma_v^2 = 0$, *i.e.*, under classical regression when the explanatory variable is observed without any measurement error. Thus θ can be regarded as a measure of departure of ultrastructural model from classical regression model.

One problem of the g-class estimator $\hat{\beta}_g$, however, is that it may have a sign which is opposite to that of b. It is straightforward to see that this will happen when $s_{xx} < (1-\frac{g}{n})\sigma_v^2$. One way to avert this undesirable property of $\hat{\beta}_g$ is to apply the procedure of the double k-class estimator (Ullah and Ullah (1978), Ohtani and Wan (2002)) on x_i's. This gives the following alternative substitute for X_i in (2.1):

$$\hat{X}_i^* = \bar{x} + \left[1-\left(1-\frac{g}{n}\right)\frac{\sigma_v^2}{s_{xx}+\frac{h}{n}\sigma_v^2}\right](x_i-\bar{x}), \tag{2.11}$$

where h is a non-negative scalar in addition to g. Now, using $\hat{X}_i^*$ in conjunction with the least squares procedure, we obtain the following family of gh-class consistent estimators:

$$\hat{\beta}_{gh} = \left[1-\left(1-\frac{g}{n}\right)\frac{\sigma_v^2}{s_{xx}+\frac{h}{n}\sigma_v^2}\right]^{-1} b \tag{2.12}$$

which is characterized by two non-negative scalars g and h. Clearly, the estimators b and $\hat{\beta}_{gh}$ will always have the same sign for positive values of h. Further, it provides a class of estimators encompassing the estimators $\hat{\beta}$ and $\hat{\beta}_g$ as special cases.

3 Large sample properties of estimators

As no specific form of the distribution of errors has been assumed, exact results on the finite sample properties of $\hat{\beta}_{gh}$ cannot be determined. Therefore, we confine our attention to its large sample asymptotic properties. For this purpose, we assume that the errors have distributions with moments of at least fourth order and

$$\lim_{n\to\infty} s_{mm} = \lim_{n\to\infty}\left[\frac{1}{n}\sum_{i=1}^{n}(m_i-\bar{m})^2\right] = \sigma_m^2, \tag{3.1}$$

which is finite, and excludes, for instance, the possibility of any trend, see Schneeweiss and Witschal (1987) and Schneeweiss (1991).

Theorem 1. *The relative bias of the estimator* $\hat{\beta}_{gh}$ *of* β *to order* $O(n^{-1})$*is given by,*

$$\begin{aligned} RB(\hat{\beta}_{gh}) &= E\left(\frac{\hat{\beta}_{gh}-\beta}{\beta}\right) \\ &= \frac{\theta}{n(1-\theta)}\left[\left(\frac{3-\theta+\theta\gamma_{2v}}{1-\theta}\right)-(g+\theta h)\right]. \end{aligned} \tag{3.2}$$

Proof. See the Appendix.

If we set $h=0$ in (3.2), then we obtain the expression of the relative bias of $\hat{\beta}_g$ given in Srivastava and Shalabh (1997a). It is also interesting to observe from (3.2) that only the kurtosis of the distribution of the measurement errors in the explanatory variable affects the bias whereas the skewness of any of the distribution plays no role at least to the order of our approximation. Furthermore, the relative bias is positive when

$$g+\theta h < \frac{3-\theta+\theta\gamma_{2v}}{1-\theta} \quad ; \quad 0 \le \theta < 1. \tag{3.3}$$

while the converse is true when (3.3) holds with a reversed inequality sign.

Now, for comparing the estimators with respect to the magnitude of bias to the given order of approximation, let us consider the square of the relative bias. It is easy to see that $\hat{\beta}_{gh}$ has a smaller magnitude of bias than $\hat{\beta} \equiv \hat{\beta}_{oo}$ when

$$g+\theta h < 2\left(\frac{3-\theta+\theta\gamma_{2v}}{1-\theta}\right) \quad ; \quad 0 \le \theta < 1. \tag{3.4}$$

Similarly, the estimator $\hat{\beta}_{gh}$ has smaller magnitude of bias compared with $\hat{\beta}_g \equiv \hat{\beta}_{g0}$ when

$$2g+\theta h < 2\left(\frac{3-\theta+\theta\gamma_{2v}}{1-\theta}\right) \quad ; \quad 0 \le \theta < 1. \tag{3.5}$$

It follows that $\hat{\beta}_{gh}$ is superior to both the estimators $\hat{\beta}$ and $\hat{\beta}_g$ with respect to the criterion of magnitude of bias, to the order of our approximation, when g and h are chosen to satisfy (3.5). As $(2+\gamma_{2v})$ is always positive and θ lies between 0 and 1, the condition (3.5) is satisfied whenever

$$2g+h < 6, \tag{3.6}$$

which provides a neat condition for the choice of the characterizing scalars g and h.

The asymptotic distribution of $\hat{\beta}_{gh}$ is normal with mean 0 and variance $E(f_{xy}^2)$ where f_{xy} is given in (4.4) in the Appendix. The expression for $E(f_{xy}^2)$ can be easily derived using the results from Srivastava and Shalabh (1997c, Appendix). It may be noted that the expression of $E(f_{xy}^2)$ is independent of the characterizing scalars g or h because f_{xy} itself is independent of g and h. The mean squared error of $\hat{\beta}_{gh}$ to order $O(n^{-1})$ is given by the expression $E(f_{xy}^2)$. In order to study the effect of g and h, we propose to study the mean squared error of $\hat{\beta}_{gh}$ to order $O(n^{-2})$ in the following Theorem.

Theorem 2. *The differences in the relative mean squared error of $\hat{\beta}_{gh}$ with $\hat{\beta}$ and $\hat{\beta}_g$, to order $O(n^{-2})$, are given by,*

$$\begin{aligned}\Delta(\hat{\beta},\ \hat{\beta}_{gh}) &= E\left(\frac{\hat{\beta}-\beta}{\beta}\right)^2 - E\left(\frac{\hat{\beta}_{gh}-\beta}{\beta}\right)^2 \\ &= \frac{\theta^2(g+\theta h)}{n^2(1-\theta)^2}\left[\frac{2(3-\theta+\theta\gamma_{2v})}{1-\theta}-(g+\theta h)\right] \end{aligned} \qquad (3.7)$$

and

$$\begin{aligned}\Delta(\hat{\beta}_g,\ \hat{\beta}_{gh}) &= E\left(\frac{\hat{\beta}_g-\beta}{\beta}\right)^2 - E\left(\frac{\hat{\beta}_{gh}-\beta}{\beta}\right)^2 \\ &= \frac{\theta^3 h}{n^2(1-\theta)^2}\left[\frac{2(3-\theta+\theta\gamma_{2v})}{1-\theta}-(2g+\theta h)\right], \end{aligned} \qquad (3.8)$$

respectively.

Proof. See the Appendix.

Once again, we observe that only the kurtosis of the distribution of the measurement errors in the explanatory variable has a pronounced effect on the efficiency of estimators. Also, it is easy to see from (3.7) that $\hat{\beta}_{gh}$ is superior to $\hat{\beta}$ in terms of relative mean squared error under the condition (3.4) which ensures smaller magnitude of bias as well. Similarly, we find from (3.8) that $\hat{\beta}_{gh}$ has not only smaller magnitude of bias but also smaller mean squared error in comparison with $\hat{\beta}_g$ when (3.5) holds true. We thus conclude that the estimator $\hat{\beta}_{gh}$ is superior to both the estimators $\hat{\beta}$ and $\hat{\beta}_g$ with respect to the twin criteria of mean squared error and the magnitude of bias at least to the order of our approximations for all kinds of distributions so long as we choose g and h to satisfy (3.6).

Now, it is observed from (3.7) that the optimal values of g and h that maximize the gain in efficiency of $\hat{\beta}_{gh}$ over $\hat{\beta}$ cannot be separately obtained. However, the optimal value of the sum $(g+\theta h)$ is given by,

$$g+\theta h = \frac{3-\theta+\theta\gamma_{2v}}{1-\theta} \qquad (3.9)$$

or

$$h = \frac{1}{\theta}\left[\frac{3-\theta+\theta\gamma_{2v}}{1-\theta}-g\right] \qquad (3.10)$$

while the choice of g is arbitrary.

Similarly, we see from (3.8) that the gain in efficiency of $\hat{\beta}_{gh}$ over $\hat{\beta}_g$ is maximized when the sum $(2g + \theta h)$ is given by

$$2g + \theta h = \frac{2(3 - \theta + \theta\gamma_{2v})}{1 - \theta}. \tag{3.11}$$

or

$$h = \frac{2}{\theta}\left[\frac{3 - \theta + \theta\gamma_{2v}}{1 - \theta} - g\right], \tag{3.12}$$

while the choice of g is arbitrary. It may be noted that the choice of h in (3.12) is twice of the choice of h in (3.10).

The above observations suggest that the largest amount of gain in efficiency of $\hat{\beta}_{gh}$ over $\hat{\beta}$ and $\hat{\beta}_g$ is achieved when g and h are chosen in such a way that (3.9) and (3.11), respectively hold true. It may also be remarked that the optimal values of $(g+\theta h)$, as specified by (3.9), and $(2g + \theta h)$, as specified by (3.11), are larger for platykurtic distributions ($\gamma_{2v} > 0$) in comparison with the optimal value for leptokurtic distribution ($\gamma_{2v} < 0$). For both cases, however, the optimal values would be substantially different from the optimal value in the case of mesokurtic distributions ($\gamma_{2v} = 0$) of which the normal distribution is a special case.

4 Appendix

Proof of Theorem 1:

Using results from Srivastava and Shalabh (1997c, Appendix), we can write

$$s_{xx} = \frac{\sigma_v^2}{\theta}\left(1 + \frac{f_{xx}}{n^{1/2}}\right) \tag{4.1}$$

and

$$s_{xy} = \beta s_{xx} - \beta\sigma_v^2\left(1 + \frac{f_{xy}}{n^{1/2}}\right), \tag{4.2}$$

where f_{xx} and f_{xy} are of order $O_p(1)$ defined as,

$$f_{xx} = \frac{\theta}{n^{1/2}\sigma_v^2}\left[\sum_{i=1}^{n}(v_i + w_i - \bar{v} - \bar{w})^2 - n(\sigma_v^2 + \sigma_w^2) + 2\sum_{i=1}^{n}(m_i - \bar{m})(v_i - w_i)\right] \tag{4.3}$$

and

$$\begin{aligned} f_{xy} \quad = \quad & \frac{\theta}{n^{1/2}\sigma_v^2}\left[\sum_{i=1}^{n}(v_i - \bar{v})^2 - n\sigma_v^2 + \sum_{i=1}^{n}(v_i - \bar{v})(m_i + w_i)\right. \\ & \left. -\frac{1}{\beta}\sum_{i=1}^{n}(u_i - \bar{u})(m_i + w_i + v_i)\right] \end{aligned} \tag{4.4}$$

respectively.

Further, we have,

$$E(f_{xy}) = -\frac{1}{n^{1/2}} \tag{4.5}$$

and

$$E(f_{xx}f_{xy}) = 2 + \theta\gamma_{2v}. \tag{4.6}$$

Using (4.3) and (4.4) in (2.12), we can write,

$$\left(\frac{\hat{\beta}_{gh} - \beta}{\beta}\right) = \left(\frac{\theta}{1-\theta}\right)\left[-\frac{f_{xy}}{n^{1/2}} + \frac{1}{n}\left[\frac{f_{xx}f_{xy}}{(1-\theta)} - (g+\theta h)\right\}\right] + O_p(n^{-3/2}), \tag{4.7}$$

whence we find,

$$E\left(\frac{\hat{\beta}_{gh} - \beta}{\beta}\right) = \left(\frac{\theta}{n(1-\theta)}\right)\left[\left(\frac{3-\theta}{1-\theta}\right) - (g+\theta h) + \left(\frac{\theta}{1-\theta}\right)\gamma_{2v}\right] \tag{4.8}$$

to order $O(n^{-1})$. This gives the result of Theorem 1.

Proof of Theorem 2:

To derive the result in Theorem 2, we observe that,

$$\begin{aligned}
\Delta(\hat{\beta}, \hat{\beta}_{gh}) &= E\left(\frac{\hat{\beta} - \beta}{\beta}\right)^2 - E\left(\frac{\hat{\beta}_{gh} - \beta}{\beta}\right)^2 \\
&= \frac{\theta^2(g+\theta h)}{n(1-\theta)^2} E\left[-\frac{2f_{xy}}{n^{1/2}} + \frac{1}{n}\left(\frac{2f_{xx}f_{xy}}{1-\theta} - (g+\theta h)\right)\right] + O(n^{-5/2}) \\
&= \frac{\theta^2(g+\theta h)}{n^2(1-\theta)^2}\left[2\left(\frac{3-\theta}{1-\theta}\right) - (g+\theta h) + \left(\frac{2\theta}{1-\theta}\right)\gamma_{2v}\right] + O(n^{-5/2}), \quad (4.9)
\end{aligned}$$

which is the result of (3.7) in Theorem 2. Equation (3.8) can be derived in a similar manner.

References

[1] Cheng, C.L. and J.W. Van Ness (1999). Statistical regression with measurement errors, Arnold Publishers, London.

[2] Dolby, G.R. (1976). The ultrastructural model: a synthesis of the functional and structural relations. *Biometrika*, **63**, *39-50.*

[3] Fuller, W.A. (1987). Measurement Error Models. Wiley, New York.

[4] Ohtani, K. and A.T.K. Wan (2002). On the use of the Stein-variance estimator in the double k-class estimators. *Econometric Reviews*, **21**, *121-134.*

[5] Saleh, A.K.Md.E (2006). Theory of preliminary test and Stein-type estimation with applications. Wiley, New York.

[6] Schneeweiss, H. (1976). Consistent estimation of a regression with errors in the variables. *Metrika*, **23**, 2 *101-115.*

[7] Schneeweiss, H. (1991). Note on a linear model with errors in the variables and with trend. *Statistical Papers*, **32**, *261-264.*

[8] Schneeweiss, H. and H. Witschel (1987). Small sample properties of estimators in a linear relationship with trend – a Monte Carlo study. In *Ökonomie und Mathematik*; Opitz, O. and B. Rauhut (eds.), Springer, Berlin, *341-352.*

[9] Shalabh (1998). Improved estimation in measurement error models through Stein-rule procedure. *Journal of Multivariate Analysis*, **67**, *35-48.*

[10] Shalabh (2000). Corrigendum of Shalabh (1998). *Journal of Multivariate Analysis*, **74**, *162.*

[11] Shalabh (2003). Consistent estimation of coefficients in measurement error models with replicated observations. *Journal of Multivariate Analysis*, **86**, 2, *227-241.*

[12] Srivastava, A.K.; Shalabh (1997a). Improved estimation of slope parameter in a linear ultrastructural model when measurement errors are not necessarily normal. *Journal of Econometrics*, **78**, *153-157.*

[13] Srivastava, A.K.; Shalabh (1997b). A new property of Stein procedure in measurement error model. *Statistics & Probability Letters*, **32**, *153-157.*

[14] Srivastava, A.K.; Shalabh (1997c). Asymptotic efficiency properties of least squares in an ultrastructural model. *TEST*, **6**, *419-431.*

[15] Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium*, University of California Press: Berkeley, Volume 1, *197-206.*

[16] Ullah, A. and S. Ullah (1978). Double k-class estimators of coefficients in linear regression. *Econometrica*, **49**, *705-722.*